Note to Students: This practice exam is intended to give you an idea of the type of questions the instructor asks and the approximate length of the exam. **It does NOT indicate the exact questions or the topics covered.** Students should refer to the appropriate learning objectives to determine the coverage of the material and appropriate questions.

NOTE: The solutions to this exam are given at the end of the exam.
Instructions:
  o Put your name, instructor, and class time in the space provided.
  o Read each question carefully.
  o Show all work to receive full credit.
  o Make all work legible.
  o Provide one answer to each question in the space provided.
  o Table packets will be provided.
  o Unless otherwise stated, use a significance level of 5% and a confidence level of 95%. For tests of hypothesis be sure to state the hypotheses, test statistic, p-value and conclusion about H₀.

Honor Pledge:

I certify that I have not received or given unauthorized aid in taking this exam.

Signed: __________________
1. (5 points) Complete the Instructor and Class Time information on the cover sheet.

True/False Multiple Choice CIRCLE the best answer (3 points each).

2. True False In a distribution that is right-skewed the mean will be greater than the median.

3. True False A t-distribution is skewed to the right.

4. True False The p-value is the probability that the null hypothesis is correct.

5. True False A correlation is always between 0 and 1.

6. True False If a list of numbers has a mean of 0, then the standard deviation will also be zero.

7. True False A test statistic is used to measure the difference between the observed sample data and what is expected when the null hypothesis is true.

8. True False An outlier will typically fall between Q1 and Q3.

9. A student takes a standardized exam. The grader reports the student’s standardized score (z-score) as –1.8. This indicates:
   a. The student scored lower than the average.
   b. The student scored less than one standard deviation from the average.
   c. A mistake has been made in calculating the score, since a standard score can never be negative.
   d. Both a and b, but not c.

10. A political scientist wants to know how college students feel about the social security system. She obtains a list of the 3114 undergraduates at her college and mails a questionnaire to 250 students selected at random. Only 100 of the questionnaires are returned. In this study, the rate of non-response would be
    a. 0.25.
    b. 0.40.
    c. 0.75.
    d. 0.60.

11. The most important advantage of experiments over observational studies is that
    a. experiments are usually easier to carry out.
    b. experiments can give better evidence of causation.
    c. the placebo effect cannot happen in experiments.
    d. an observational study cannot have a response variable.

12. We use the t-distribution to calculate a confidence interval for the population mean \( \mu \). If we increase the sample size from 10 to 20 the interval would become smaller because of
    a. the change in degrees of freedom.
    b. the change in standard error.
    c. both a and b.
    d. none of the above.
13. We would like to test the hypothesis that $\mu=20$ versus the alternative that $\mu\neq20$. From a sample of 30 subjects we calculate the test statistic to be $t=2.3$. The p-value would be
   a. 29.
   b. 0.014.
   c. 0.986.
   d. 0.028.

14. A correlation of $r=0.85$ indicates that the graph of the data would show
   a. Points tightly packed around a line that slopes up to the right.
   b. Points tightly packed around a line that slopes down to the right.
   c. Points widely scattered around a line that slopes up to the right.
   d. Points widely scattered around a line that slopes down to the left.

15. We conduct a regression and find that the least squares line is $y=3+5x$. This indicates that as the value of $x$ increases by 4 the expected value of $y$ would increase by:
   a. 5.
   b. 8.
   c. 23.
   d. 20.

16. A study is conducted to examine the impact of a new medicine on the cholesterol level in adult males. A group of 20 subjects have their blood tested to examine their cholesterol level. They are then given the medicine for one week. At the end of the week their cholesterol is again measured. The best way to determine if their was a significant change in blood cholesterol we should
   a. Create a histogram of the 20 cholesterol levels.
   b. Create side by side box plots of the cholesterol level before and after the week.
   c. Calculate the regression to predict the cholesterol after using the cholesterol before.
   d. Conduct a paired difference t-test for the change in cholesterol.
A study was conducted on the amount of time drivers wait for a stoplight to change at a particular intersection. The amount of time spent by 300 drivers was recorded and the resulting data were used to create this boxplot.

17. The median amount of time spent at this traffic light was
   a. 1.0.
   b. 2.3.
   c. 4.0.
   d. It is impossible to tell without the standard deviation.

18. The top 25% of drivers waited over
   a. 1.3.
   b. 2.3.
   c. 4.0.
   d. It is impossible to tell without the standard deviation.

19. The mean amount of time spent at this traffic light was
   a. greater than the median.
   b. less than the median.
   c. about the same as the median.
   d. It is impossible to tell without the standard deviation.
The midterm exam grades of a history course were used to create the following stem and leaf plot.

```
Stem-and-leaf of visitors  N  =  40
Leaf Unit = 1.0

2  5
3
4
5  26
6  1359
7  00338
8  01122335789
9  0011123345677899
```

20. What is the first quartile of these scores?
   a. 52
   b. 73
   c. 86
   d. 93

21. The distribution of these exam scores would best be described as
   a. skewed to the right.
   b. skewed to the left.
   c. bimodal.
   d. symmetric.

22. The instructor was concerned that the highest score was only 99 (instead of 100)
He decided to add one point to everyone’s score. The effect of this would be
   a. The standard deviation would increase by 1.
   b. The median would change but the mean would not since this is a skewed distribution.
   c. The mean would change but the median would not since this is a skewed distribution.
   d. The standard deviation would not change but the mean and median would increase.

23. The sampling distribution of a statistic is
   a. the probability that we obtain the statistic in repeated random samples of the same size from the same population.
   b. the mechanism that determines whether randomization was effective.
   c. the extent to which the sample results differ systematically from the truth.
   d. the distribution of values taken by a statistic in all possible samples of the same size from the same population.
Does the type of movie children are watching make a difference in the amount of
snacks they will eat? A group of 50 children were randomly assigned to watch either
a cartoon or a live action musical (25 to each). Crackers were available in a bowl,
and the investigators compared the number of crackers eaten by children while
watching the different kinds of movies.

24. In this study the explanatory variable is:
   a. The amount of crackers eaten.
   b. The children.
   c. Does the type of movie make a difference in the amount of snacks eaten?
   d. The type of movie watched.

25. This study is best described as:
   a. A placebo controlled experiment.
   b. A matched pairs experiment.
   c. A randomized observational study.
   d. A randomized comparative experiment.

26. In the study described above, one kind of movie was shown at 8 AM (right after
the children had breakfast) and another at 11 AM (right before the children had
lunch). It was found that during the movie shown at 11 AM, more crackers were
eaten than during the movie shown at 8 AM. The investigators concluded that the
different types of movies had an effect on appetite. The results cannot be trusted
because
   a. the study was not double blind. Neither the investigators nor the children
      should have been aware of which movie was being shown.
   b. the investigators were biased. They knew beforehand what they hoped the
      study would show.
   c. the investigators should have used several bowls, with crackers randomly
      placed in each.
   d. the time the movie was shown is a confounding variable.

27. An educational software company wants to assess the usefulness of its software. It
runs a “quick vote” poll on a website, asking users to indicate whether they like or
dislike the software. Of 900 respondents, 610 said they liked the software. The
results of the sample are probably
   a. unbiased, because of the large sample size.
   b. biased, because it is a voluntary response sample.
   c. unbiased, because it is a simple random sample.
   d. unbiased, because a larger sample should be used.
28. Match the summary statistics with the histograms (3 points each).
   a. mean = 4.99, median = 3.13, standard deviation = 5.49 ________
   b. mean = 4.89, median = 4.83, standard deviation = 7.99 ________
   c. mean = 5.01, median = 6.87, standard deviation = 5.49_________
   d. mean = 4.96, median = 4.93, standard deviation = 0.96_________

29. (5 points) The histogram for Variable 3 above displays the distribution of the amount of money spent on lunch during a six-day period by a group of students in an English class. Suppose every one of these students attended a lunch that cost them $3.00 on the seventh day. What would be the mean, median, and standard deviation of the amount they spent on lunch for the whole one-week period?

   Mean:_______
   Median:______
   SD:_________
30. (15 points) A study was conducted to examine the impact of speaking in public on college students. A class of 15 statistics students participated in the study. At the beginning of a lecture the students recorded their systolic blood pressure. During the lecture the instructor called on each student to stand and answer questions about topics in the lecture. After speaking the students once again recorded their blood pressure. The resulting values are given below along with summary statistics.

<table>
<thead>
<tr>
<th>Before</th>
<th>102</th>
<th>123</th>
<th>128</th>
<th>125</th>
<th>136</th>
<th>108</th>
<th>130</th>
<th>113</th>
<th>130</th>
<th>136</th>
<th>146</th>
<th>138</th>
<th>108</th>
<th>149</th>
<th>132</th>
</tr>
</thead>
<tbody>
<tr>
<td>After</td>
<td>141</td>
<td>142</td>
<td>146</td>
<td>136</td>
<td>144</td>
<td>111</td>
<td>132</td>
<td>114</td>
<td>103</td>
<td>133</td>
<td>140</td>
<td>130</td>
<td>100</td>
<td>133</td>
<td>113</td>
</tr>
<tr>
<td>Difference (Before-After)</td>
<td>-39</td>
<td>-19</td>
<td>-18</td>
<td>-11</td>
<td>-8</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Before</th>
<th>Mean = 125.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>After</td>
<td>Mean = 127.9</td>
</tr>
<tr>
<td>StDev</td>
<td>StDev = 15.03</td>
</tr>
<tr>
<td>StDev</td>
<td>StDev = 15.5</td>
</tr>
<tr>
<td>StDev</td>
<td>StDev = 14.87</td>
</tr>
</tbody>
</table>

a. Does this information indicate that blood pressure increased because of speaking? Conduct an appropriate hypothesis test and show appropriate steps.

\[ H_0: \] ______________
\[ H_a: \] ______________

Test Statistic: ________

p-value: ________

Conclusion (circle one): Reject \( H_0 \)  Do not reject \( H_0 \)

b. What does your conclusion about \( H_0 \) tell us about blood pressure and the answering questions in class? Explain.
31. (10 points) In statistics we are able to make inference because statistics have a predictable distribution called a sampling distribution. One method of inference we have discussed is the idea of hypothesis testing. Explain briefly how the sampling distribution is used in the process of hypothesis testing.

32. Customers using a self-service soda dispenser take an average of 12 ounces of soda with an SD of 4 ounces. Assume that the amount would be normally distributed.
   a. (5 points) What is the chance that a randomly selected customer takes over 10 ounces of soda?

Answer:_______

b. (5 points) What is the chance that a randomly selected customer takes between 13 and 14 ounces of soda?

Answer:_______

c. (5 points) What is the chance that the next 100 customers will take an average of less than 12.24 ounces?

Answer:_______
33. A consumer advocacy group recorded several variables on 140 models of cars. The resulting information was used to produce the following regression output that relates the city gas mileage (in mpg) and the engine displacement (in cubic inches).

The regression equation is

\[ \text{mpg:city} = 33.4 - 0.0624 \text{displacement} \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>33.4</td>
<td>0.7762</td>
<td>43.00</td>
<td>0.000</td>
</tr>
<tr>
<td>displacement</td>
<td>-0.0624</td>
<td>0.003810</td>
<td>_____</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ S = 3.13923 \quad R-Sq = 66.0\% \quad R-Sq(adj) = 65.8\% \]

a. (5 points) We have a car that has an engine with 150 cubic inches. Based on this output, what city gas mileage would you predict for this car?

Answer:_______

b. (5 points) Based on this output what is the correlation between city gas mileage and displacement?

Answer:_______

c. (5 points) The test statistics for testing the slope is zero is missing. Calculate this value:

Answer:_______

The group also recorded the power of the engine (in horsepower) for each car. The following regression output was produced.

The regression equation is

\[ \text{mpg:city} = 32.2 - 0.0572 \text{horsepower} \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>32.2</td>
<td>1.000</td>
<td>32.19</td>
<td>0.000</td>
</tr>
<tr>
<td>horsepower</td>
<td>-0.0572</td>
<td>0.005022</td>
<td>-11.39</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ S = 3.86688 \quad R-Sq = 48.4\% \quad R-Sq(adj) = 48.1\% \]

d. (5 points) Which of the variables (horsepower or displacement) would be the better predictor of city gas mileage? Explain.
34. A large university is planning to build a new parking garage on campus that will be paid for by a larger student parking fee. The student newspaper wanted to determine what proportion of the students favor the new garage. They email a random sample of 100 students and find that 30 of them are in favor of the new garage.
   a. (5 points) What is the parameter of interest in this study?
   b. (10 points) Calculate a 99% confidence interval for the parameter of interest in this situation.

35. (15 points) A genetic theory says that a cross between two pink flowering plants will produce red flowering plants 25% of the time. To test the theory, 100 crosses are made and 31 of them produce a red flowering plant. Is this strong evidence that the theory is wrong? Carry out the appropriate hypothesis test. Be sure to write down the null and alternative hypotheses, find the test statistic and the P-value, and state your conclusions.

   Test Statistic: __________
   p-value: ______________

   Versions 1a Page 12 of 17
36. A researcher is going to conduct a regression and created the following scatter plot that relates the variable $y$ to the variable $x$.

(5 points) Explain why it might not be a good idea to make inference using regression in this situation? Explain and list what assumption would be inappropriate?
SOLUTIONS: (Learning Objectives covered are included within parentheses).

1. Instructor and Course Information
2. (C2) True
3. (F10) False
4. (G3) False
5. (J4) False
6. (C3) False
7. (G5) True
8. (B9) False
9. (D4) A
10. (A3) D
11. (L1) B
12. (F13) C
13. (G4) D
14. (J7) A
15. (J9) D
16. (H1) D
17. (C8) B
18. (C8) C
19. (C2) A
20. (C1) B
21. (B6) B
22. (C4) D
23. (E1) D
24. (J1) D
25. (L1) D
26. (L3) D
27. (A3) B
28. (C6) (a) 3  (b) 2  (c) 4  (d) 1

29. (C4) Mean 7.99  Median 6.13  SD 5.49
30. (H3)

(a) 
Hypotheses: 
\[ H_0 : \mu = 0 \]
\[ H_A : \mu < 0 \]

Test Statistic: 
\[ t_{15-1-14} = \frac{-2.6 - 0}{14.87 / \sqrt{15}} = -.677186 \]

The p-value is greater than .10 since the p-value for \( t_{14} = -1.35 \) is .10.
Do not reject \( H_0 \).

(b) From our sample data, there is no evidence at a \( \alpha = .05 \) level that blood pressure increased because of speaking.

31. (E1) Statistics have a predictable distribution called a sampling distribution. The sampling distribution allows us to quantify the variability in sample statistics. This allows us to calculate a p-value, which is the probability of observing a test statistic that is as extreme or more extreme than our test statistic assuming the null hypothesis is true.

32. (D4,D5) Let \( X \) be the number of ounces taken from a self-service soda dispenser.

Then

(a) 
\[ P(X > 10) = P\left( \frac{X - 12}{4} > \frac{10 - 12}{4} \right) \]
\[ = P(Z > -0.5) = 1 - P(Z < -0.5) \]
\[ = 1 - .3085 = .6915 \]

(b) 
\[ P(13 < X < 14) = P\left( \frac{13 - 12}{4} < \frac{X - 12}{4} < \frac{14 - 12}{4} \right) \]
\[ = P(0.25 < Z < 0.5) \]
\[ = P(Z < 0.5) - P(Z < 0.25) \]
\[ = .6915 - .5987 = .0928 \]
(c) 
\[ P\left( \bar{X} < 12.24 \right) = P\left( \frac{\bar{X} - 12}{4/\sqrt{100}} < \frac{12.24 - 12}{4/\sqrt{100}} \right) = P\left( Z < 0.6 \right) = .7257 \]

33. (a) (J9) \( \hat{y} = 33.4 - 0.0624(150) = 24.04 \)

(b) (J8) \( r = -\sqrt{.66} = -.8124 \), negative value for the correlation since the slope is negative.

(c) (K4) \( t = \frac{-0.0624 - 0}{0.00381} = -16.38 \)

(d) (J13) Displacement, more variability in the city gas mileage (66%) is explained by displacement than horsepower, which explains 48.4% of the variability in city gas mileage.

34. (a) (A1) The population proportion of university students in favor of the new parking garage.

(b) (F5) 
\[ \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]
\[ \Rightarrow 0.3 \pm 2.576 \sqrt{\frac{0.3(1-0.3)}{100}} \]
\[ \Rightarrow 0.3 \pm 2.576(0.0458) \]
\[ \Rightarrow 0.3 \pm 0.118 \]
\[ \Rightarrow (0.182, 0.418) \]

35. (g6) 
Hypotheses: 
\[ H_0 : p = 0.25 \]
\[ H_A : p \neq 0.25 \]

Test Statistic: 
\[ z = \frac{0.31 - 0.25}{\sqrt{0.25(1-0.25)/100}} = \frac{0.06}{0.04330127} = 1.3856 \approx 1.39 \]

p-value = \( 2P(Z > 1.39) = 2(1 - P(Z < 1.39)) = 2(1 - .9177) = .1646 \)
The p-value for the test is greater than the $\alpha = .05$ level, therefore we fail to reject the null hypothesis. There is no evidence at a $\alpha = .05$ that a cross between two pink flowering plants will be different from 25% red flowering plants.

Our data does support the theory that a cross between two pink flowering plants will produce red flowering plants 25% of the time.

36. (J10) Linear regression may not be appropriate in this situation because the scatter plot of the data indicate a curved relationship between X and Y. The data do not appear to have a linear relationship so linear regression is not appropriate.